A Multilevel Convolutional Recurrent Neural Network for Blade Icing Detection of Wind Turbine

Weiwei Tian, Xu Cheng, *Member, IEEE*, Guoyuan Li, *Senior Member, IEEE*, Fan Shi, *Member, IEEE*, Shengyong Chen, *Senior Member, IEEE*, and Houxiang Zhang, *Senior Member, IEEE*

***Abstract*—Blade icing detection becomes increasingly significant as it can avoid revenue loss and power degradation. Conventional methods are usually limited by additional costs, and model-driven methods heavily depend on prior domain knowledge. Data-driven methods, especially deep learning approaches without needing the time-consuming handcraft feature engineering, offer a promising solution for blade icing detection. However, the monitoring signals normally have complex and diverse features as wind turbine operates in complex environments, thus effective model is needed for data analyzing. Additionally, the distribution of monitoring data is imbalanced, which causes the abnormal data mining inadequate. In this work, a multilevel convolutional recurrent neural network (MCRNN), is proposed for blade icing detection. Specifically, discrete wavelet decomposition is leveraged to obtain multilevel features both from the time domain and the frequency domain. A parallel structure combining an LSTM branch and a CNN branch is established in each level for feature extraction. To alleviate the severe data imbalance, two mechanisms, including data resampling algorithm and class-rebalanced loss function, are investigated. Furthermore, a multi-step accumulation strategy is proposed to enhance the accuracy of real-time detection. Extensive studies demonstrate that the proposed MCRNN can achieve up 38.8% and 42.9% higher F1-score over the best baseline on the balanced data sets processed by data resampling algorithm and 23.9% and 30.6% higher on imbalanced data sets with MCRNN optimized by the class-rebalanced loss function. The real-time detection verifies the applicability of the proposed method and indicates that the proposed multi-step accumulation strategy can improve the accuracy of icing detection.**

***Index Terms*—Wind turbine, Icing detection, Deep learning, Discrete wavelet decomposition, Neural network.**

I. Introduction

RENEWABLE energy sources have gained much attention due to the global energy crisis and the increasing demand for clean energy [1]. Because of its characteristics of abundant availability, cost-effectiveness, and technological maturity for commercial use, wind power is currently one of the fastest-growing renewable energy segments worldwide [2]. However, blade icing is a strong limitation for the performance of wind turbines in cold climates. Wind farms are established mostly based on high altitude mountainous areas with low temperature and high humidity, where wind turbine blades are susceptible to be icing, especially in winter. Ice accretion on blades not only affects the power generation by wind turbines with up to 30% loss in annual power generation in severe cases, but also gives rise to safety problems in the vicinity of the wind power plants [3]. Therefore, icing detection of wind turbine blade has been receiving increasing attention. To minimize losses caused by blade icing, to date, much effort have been invested in icing condition detection.

Conventional methods for blade icing detection of wind turbine can be simply divided into direct and indirect techniques. The direct methods are based on the physical properties of ice, while the indirect methods are determined by measuring parameters, and these parameters are affected by the accretion of ice [4]. Direct detection methods mainly include ultrasonic damping [5], resonance frequency measurement [6], optical measurement technology [7] and so on. Indirect detection techniques include comparison of output power [3], heated and unheated anemometers [8], dew point and temperature [6], etc. The advantage of the conventional methods is that highly reliable measurement (i.e. the thickness and presence of ice [9], [3]) can be observed. However, the disadvantage is obvious that these technologies may bring negative affect of blade structure and are limited by their high costs and huge personnel requirements, or are inaccurate in icing detection and insensitive to small amounts of ice [10].

In order to overcome the shortcomings of conventional approaches, recently, blade icing detection methods based on monitoring signals of wind turbines have received extensive attention due to their lower cost and less mechanical changes. These methods consist of model-driven methods and data-driven methods. The model-driven methods employ domain knowledge to build a mathematical model to reflect the relationship between monitoring signals and icing. A mathematical model was proposed by Corradini et al. for blade icing monitoring of wind turbines based on an observer to estimate error overpasses a suitable threshold of the rotor angular speed [11]. Shi introduced a model-based blade icing detection method, which does not require additional measurements and can be used for any type of blade aerodynamic changes, not just ice on the blade [12]. An icing model was developed by Hu et al. to better understand the icing condition of the blade. The model was verified by numerical simulation of the ice on the rotating NREL Phase VI blades [13]. However, it is challenging to obtain an accurate model for the blade icing estimation of wind turbine. Because the model-driven methods are heavily relying on the prior domain knowledge and some assumptions. In addition, external tools, such as wind tunnel,

Weiwei Tian, Xu Cheng, Fan Shi and Shengyong Chen are with the School of Computer Science and Technology, Tianjin University of Technology, Tianjin, 300384, China.

Guoyuan Li, and Houxiang Zhang are with the Department of Ocean Operations and Civil Engineering, Norwegian University of Science and Technology, Aalesund, 6009 Norway.

Weiwei Tian and Xu Cheng are equal contribution. Corresponding author: Guoyuan Li, e-mail:guoyuan.li@ntnu.no.

may be required to tune the model parameters.

By contrast, data-driven methods utilize shallow machine learning and deep learning approaches for icing detection of wind turbine blades. Shallow machine learning methods such as linear discriminant analysis were taken to indicate features of power and wind speed distribution, then multiple random forest classifiers was utilized for icing detection in [14]. Besides, Georgios employed kNN to perform icing failure detection according to the correlation of wind speed and power [15]. Support vector machine (SVM) is also used for wind turbine icing detection in [16]. Yi et al. proposed an imbalanced classification model for blade icing detection of wind turbine. A novel minority clustering SMOTE (synthetic minority oversampling technique) method is presented to overcome the imbalance in the collected sensor data [17]. These shallow machine learning methods do not rely on precise mathematical models. Nevertheless, much effort on hand-crafted feature design must be exerted [18].

Deep learning methods do not need to manually extract features [19]. It can automatically learn high-level feature representations of monitoring sensor data, which has shown potential for application of blade icing detection of wind turbines. Chen et al. proposed a deep neural network to learn discriminative features for the model construction. The phenomenon of data imbalance for the blade icing detection is clearly figured out and a triplet loss function is employed to preserve intra-class and inter-class information [20]. Liu et al. proposed a deep neural network which combines deep learning and ensemble learning technique for the improvement of model accuracy and generalization capability [21]. Cheng et al. proposed a temporal attention convolutional neural network for the blade icing detection [22]. Yuan et al. [23] studied blade wind turbines icing detection combining the wavelet transformation with fully convolutional neural network (FCNN). However, the time dependency is not considered in the proposed waveletFCNN, which may not fit in the actual application scenarios. Despite the successful application of deep learning in other domain, such as computer vision and natural language processing, it is still lacking of in-depth investigation of the applications in blade icing detection of wind turbine. There are still some challenges for the use of deep-learning-based methods for detection of the icing of wind turbine blades.

1) Data aspect: The wind turbine works in the normal state most of the time, while the time when the blades are icing is relatively less, which cause the data in normal state is much more than that in abnormal state, so the monitoring data collected by the supervisory control and data acquisition (SCADA) system is imbalanced between classes. Using these imbalanced data for icing detection model training may result in biased identification of icing in the blade.
2) Incomplete information: Data collected from SCADA system is inevitably subject to information loss from sensor damage. General treatments for missing data include imputation, mean substitution and etc. However, the detection task suffers from accumulated errors by only applying data completion algorithms on the incomplete data.
3) Model aspect: Wind turbine is a complex engineering system with fickle working circumstance, which results in the monitoring signals are apt to have complex and diverse features. Therefore, learning useful features by deep learning methods from the complex multivariate signals to identify blade icing is challenging.
4) Application aspect. The data-driven models are usually trained in an offline fashion and then are deployed for the real-time estimation. Although the high accuracy data driven model can be achieved during the offline training and testing stage, the real world situation is more complicated which may cause the high-accuracy offline trained model with a high error estimation rate or false alarm rate in the model deployment of application scenarios. Therefore, high-availability methods should be explored for actual scenarios.

To address the aforementioned challenges, a multilevel convolutional recurrent neural network (MCRNN) is proposed. Specifically, for the data aspect, data analysis and preprocessing are conducted to reduce the influence of noise and redundant information. To alleviate the severe data imbalance, two solutions, including data resampling algorithm and class-rebalanced loss function, are investigated. For incomplete information, different methods are used for data interpolation according to the duration of the missing value intervals, including linear interpolation algorithm and direct interpolation using data of the same period in different year. For the model aspect, multilevel discrete wavelet decomposition (MDWD) is first conducted on the original data to obtain wavelet detail coefficients that reveal the signal variance in multiple scales to disclose the information in both the time and frequency domains. Then, a parallel structure combining an LSTM branch and a CNN branch is then established in each level to learn knowledge from the correlation of multivariate signals. Finally, to achieve high availability in real scenarios, a multi-step accumulation strategy is introduced to reduce error estimation rate or false alarm rate in the model deployment of application scenarios.

The main contributions of this work are:

1) A novel deep neural network MCRNN is proposed for blade icing detection of wind turbine. The MDWD is first utilized to obtain multilevel features. A parallel structure network is then employed on each multilevel feature to capture the complicated correlations in both frequency and temporal domains. A data re-sampling method and a loss function based method are investigated to alleviate the imbalance of the monitoring data. A multi-step accumulation strategy is proposed to effectively reduce the false alarm rate in the real-time detection phase.
2) The proposed MCRNN is evaluated on real-world datasets from two wind turbines. Based on the baseline comparison, the proposed MCRNN demonstrates its effectiveness and superior performance. The proposed MCRNN is further demonstrated and compared with several model variants. Sensitivity analysis on MCRNN determines the optimal hyperparameters. Real-time prediction verifies the applicability of the method.

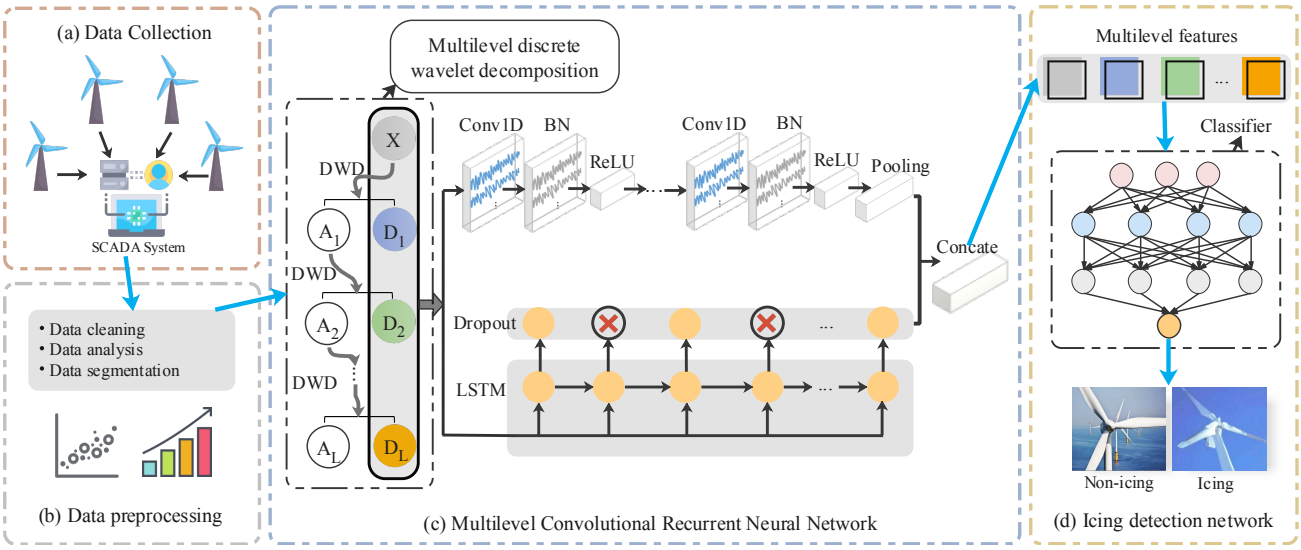

Fig. 1. Overall structure of the proposed method. (a) The SCADA system collects wind turbine operating data. (b) Analysis and preprocessing of the raw data. (c) The processed data sets are input to the proposed MCRNN model for feature extraction and knowledge learning. X is the input sensor data, A_l and D_l represent the approximate coefficient and detail coefficient of each level in multilevel discrete wavelet decomposition. BN and ReLU indicate the batch normalization layer and the ReLU layer, respectively. (d) Feature classification and icing detection.

The rest of the paper is organized as follows: The overall architecture of the proposed MCRNN is introduced in Section II. Section III verifies the effectiveness and superiority of the proposed model. Section IV summarizes the whole paper.

II. Multilevel Convolutional Recurrent Neural Network for Icing Detection

This section introduces the overall architecture of the proposed multilevel convolutional recurrent neural network for icing detection. The whole workflow is first introduced. Then, each part is described in detail.

A. Workflow structure

As shown in Fig. 1, the overall architecture of the proposed multilevel convolutional recurrent neural network for icing detection contains data collection and processing, feature extraction and knowledge learning by MCRNN and icing detection. The collected sensor data is inevitably contaminated for various reasons, and therefore it is necessary to clean the data in order to minimize the effect of noise. Furthermore, to address the challenge of the data aspect mentioned in Section I, the imbalanced data are also analyzed and processed. The processed sensor data is then input to MCRNN for feature extraction and knowledge learning. The details are illustrated in Section II-C. The features extracted by MCRNN are finally input to a designed classifier for icing detection and the optimal models for are employed for real-time icing detection.

B. Data collection and processing

The data used in this paper is collected from the SCADA system that needs to use the existing sensors only. The data mainly include the data regarding the status of the wind turbine components, and internal and external environmental conditions. There are four steps for data processing in this paper, namely, data labeling, visualization, imbalanced data processing, and data split and normalization.

1) Data labeling. The raw data are labeled by experienced engineers. Due to the unstable operation of wind turbines, some uncertain intervals are removed. Uncertain intervals are those that are very difficult to label as either normal or icing state even for these experienced engineers.

2) Missing value processing. The missing data problem of monitoring signals is ubiquitous due to various factors such as sensor damage and human error. If a single data point or several data points are lost, the linear interpolation algorithm is applied to interpolate the missing data points. When the data loss intervals last for several hours or even a single day, the operation state of wind turbine may have changed for several times, thus leveraging the interpolation algorithm to complete the data sets will bring larger error. So in this case, data at the same time in different year are used for data interpolation.

3) Statistics. To fully characterize the features of the labeled data, statistics methods can be useful for the understanding of the sensor data. Statistics has the following two beneficial aspects: i) Identifying the redundant information that makes little or no contribution to the detection of blade icing. These redundant data will be discarded directly because they may lead to poor feature representation ability. ii) Visualization of the correlation between signals. If there exists high correlation between two or more signals, it will lead to a severe degree of model overfitting. To alleviate the overfitting of the model, the average value of highly correlated variables is used as a new variable replacing the highly correlated variables for the icing detection carried out in this paper.

4) Imbalanced data processing. Class imbalance is a very common problem in the real world [24]. The data collected from SCADA system show severe imbalance because the wind turbines usually work in normal conditions most of the time, and the blades will freeze only a few times. If the trained

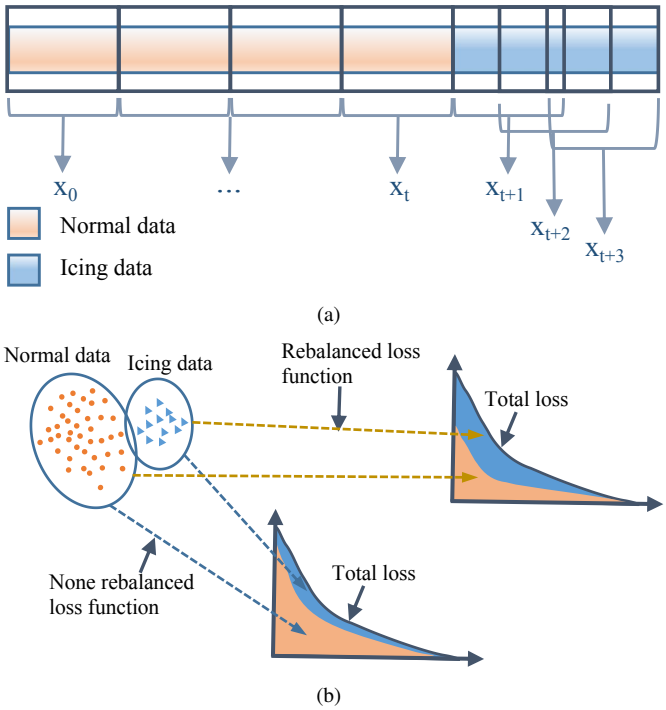

Fig. 2. Two solutions for imbalanced data processing. (a) Resampling is performed during data segmentation to generate balanced datasets. (b) A rebalanced loss function is used to optimize the model.

model cannot handle the data imbalance well, the prediction results of the model will be biased. To address this challenge, two solutions, including data resampling algorithm and class-rebalanced loss function, are explored. As shown in Fig. 2 (a), a data resampling based method is utilized to generate balanced datasets. For the normal data, the samples are obtained by utilizing the sliding window without overlapping. While, for the icing data, we can get the same number of samples by using the sliding window with overlapping. For example, in Fig. 2 (a), we can get four samples for normal data, and there are also four samples for icing data by using the proposed data resampling approach. Otherwise, there are only two samples using the sliding window without overlapping. As depicted in Fig. 2 (b), a class-rebalanced loss function is leveraged to optimize the model by assigning more weights to the icing data (minor class).

5) Data split and normalization. The output data of the SCADA system comprise a discrete sequence because the SCADA system fetches signals from different sensors approximately every 7 seconds. To ensure that each sample input to the MCRNN has a certain time range, it is necessary to split the original data with a fixed step size. In addition, data normalization is also a highly important step for data processing. The data normalization can be helpful to remove the unit limitation of the data and convert it into a dimensionless pure value, which is convenient for enabling the simultaneous analysis of the indicators with different units or magnitudes.

C. Multilevel convolutional recurrent neural network

The proposed multilevel convolutional recurrent neural network consists of three parts: discrete wavelet decomposition, LSTM branch and CNN branch. Specifically, as illustrated in Fig. 1 (c), the raw monitoring signals are first transformed by MDWD for multilevel features. Then, the original signals along with the discrete wavelet decomposition coefficients in each level are taken by the LSTM branch and CNN branch separately. The LSTM branch is used for temporal features learning. The CNN branch is used for capturing spatial information.

1) Multilevel discrete wavelet decomposition: Wind turbine is a complex engineering system with fickle working circumstance, which results in the monitoring signals acquired by the SCADA system are apt to have complex and diverse features. Wavelet decomposition is well-known method for the analysis of a time series because they can capture the features both from the time and frequency domains [25]. It is advantageous to transform the signals by MDWD to obtain further serviceable information.

The time series can be decomposed by MDWD into groups of multilevel sub-series which are ranked in order of frequency from high to low. These frequencies are beneficial and crucial for feature learning in frequency domain. We denote a sequence segment of each channel as $x_c = [x_0, \cdots, x_n, \cdots, x_{N-1}]$, $c = 1, 2, \cdots, C$, where C is the total channel number of the data. N is the length of sequences, and $[x_1, \cdots, x_c, \cdots, x_C] \in R^{N \times C}$. According to Fig. 1 (c), the 1-dimensional discrete wavelet transform is conducted on each input signal channel x_c until a specific level L. Continuous decomposition is only conducted on low-frequency components that generated from the 1/2 down-sampling of the intermediate variable sequences. In each level, the sequences represented by A_l are the approximate coefficients of the signal, which is the low-frequency component generated by applying a low pass filter $\phi = [\phi_0, \cdots, \phi_m, \cdots, \phi_M]$, and D_l are the detail coefficients of the signal, which is the high-frequency component generated by a high pass filter $\varphi = [\varphi_0, \cdots, \varphi_m, \cdots, \varphi_M]$, and $M << N$. The subseries of the upper level are convoluted as:

$$A_{l+1}[k] = \sum_{m=0}^{M} x_l[k+m] \cdot \phi_{[m]} \tag{1}$$

$$D_{l+1}[k] = \sum_{m=0}^{M} x_l[k+m] \cdot \varphi_{[m]} \tag{2}$$

where A_{l+1} and D_{l+1} denote the subapproximate coefficients and subdetail coefficients, respectively. x_0 is the input series, and $x_l[k]$ is the k-th element of the low-frequency components in the l th level, $l = 0, 1, \cdots, L$. The approximation coefficients represent some smoothed averages of the input signal, thus, we only augment the detail coefficients to the original sequences because the approximation coefficients may be a portion of the redundant data that hinder model training. Subsequently, the augmented segment for each channel can be represented by $X_c = [x_c, d_c^1, \cdots, d_c^L]$.

2) LSTM branch: Monitoring data of wind turbine is presented in the form of time series which is not independent data but a series of discrete data with temporal dependence. It is critical for the proposed model to extract potential and

valuable information from the original sequences with the transformed sequences together. The leverage of LSTM is based on the insight that the temporal correlations of the points hidden in a time series are closely related to frequency [26]. Besides, LSTM is suitable for processing and predicting important events with long time dependencies in time series. Moreover, Dropout layer is concatenated with the LSTM layer to improve the generalization capability. According to the previous subsection, we define the input of the LSTM branch as: $X_{in} = [X, D_1, \cdots, D_L] = [[x_0^t], \cdots, [x_l^t], \cdots, [x_L^t]]$, where $X_l \in R^{C \times \frac{N}{2^l}}$, and t indicates the time stamps of each subsequence. The computation at each time step of the LSTM layer is described by Graves et al. [27], and to be consistent with the notations we used, we reformulate the definition as:

$$\begin{aligned} g^i &= \sigma(W^i h^{t-1} + I^i x_l^t) \\ g^f &= \sigma(W^f h^{t-1} + I^f x_l^t) \\ g^o &= \sigma(W^o h^{t-1} + I^o x_l^t) \\ g^c &= tanh(W^c h^{t-1} + I^c x_l^t) \\ m^t &= g^f \otimes m^{t-1} + g^i \otimes g^c \\ h^t &= tanh(g^o \otimes m^t) \\ y &= f_{dropout,p}(g^o) \end{aligned} \tag{3}$$

where g^i, g^f, g^o, g^c are the activation vectors of the input, forget, output and cell state gates respectively, and σ is the logistic sigmoid function. The recurrent weight matrices are described by W^i, W^f, W^o, W^c, and I^i, I^f, I^o, I^c jointly represent the projection matrices. The hidden state vector is expressed as h^t. Simultaneously, $\otimes$ is defined as elementwise multiplication. Finally, the calculation process of the dropout is defined as $f_{dropout,p}$, where p is the probability of retaining a unit in the network.

3) CNN branch: CNN is utilized as it can capture compact data features of the short time contextual information in time series. The CNN branch consists of basic convolution neural network blocks composed of a convolutional layer followed by the batch normalization layer [28] and the ReLU layer [29]. The convolution layer can effectively fuse temporal information and channel information, so that the crucial feature of raw data and detail coefficients for icing detection can be completely extracted. The batch normalization (BN) layer is added to accelerate the training and improve generalization. The ReLU layer can alleviate the overfitting of the model because the interdependence of the parameters will be reduced. The input of the CNN branch is the same as the LSTM branch. Besides, we define X_{in} in the LSTM branch as X_0 in the CNN branch. The basic convolution block is formulated as:

$$X_i = ReLU(BN(f(W * X_{i-1} + b))), i = 0, 1, \cdots, I. \tag{4}$$

where X_i is the input or the output of each basic convolutional block and I is the total number of the basic blocks. f is the activation function, and $*$ represent the convolutional operation. Meanwhile, the parameters learned by convolutional layer are defined as W and b. Then, following the basic convolution neural network blocks, a global average pooling layer is employed to achieved dimensionality reduction and reduce the number of the parameters of the network. The outcomes of the LSTM branch and the CNN branch and the ultimate outcomes of the LSTM-CNN are merged by the concatenate layer.

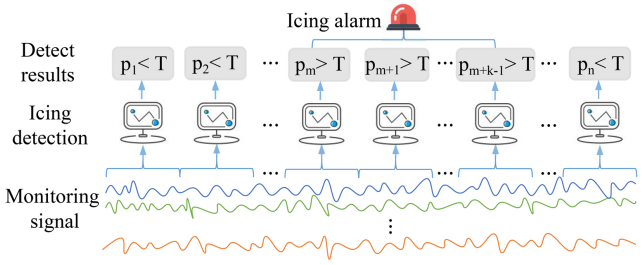

Fig. 3. An illustration of the multi-step accumulation strategy. p_n is the icing possibility and T is the threshold.

D. Icing detection

1) Icing detection classifier: Multilevel features learned by MCRNN, as illustrated in Fig. 1 (d), are input to a classifier for blades icing detection. The classifier is employed for calculating the classification probability distribution that indicates the probability of the true category (blade icing), so that the classifier can generate a prediction of whether or not icing is present on the blades. In this work, a dense layer associated with softmax function is employed as the classifier. The dense layer multiplies the weight matrix by the input vector and adds a bias to map n real numbers belonging to $(-\infty, +\infty)$ to K real numbers (fractions) belonging to $(-\infty, +\infty)$; K real numbers belonging to $(-\infty, +\infty)$ are mapped to K numbers (probability) belonging to (0, 1) by the softmax function, while ensuring that the sum is 1, where K is the total number of categories. The complete process of the dense layer along with softmax function can be summarized as:

$$\hat{y} = softmax(z) = softmax(W^T x + b) \tag{5}$$

$$softmax(z_j) = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}, j = 1, \cdots, K. \tag{6}$$

where x is the input of the dense layer, W and b represent the weight and the bias term respectively, and $\hat{y}$ is the probability of the classes. Training sets and validation sets are used to train the best models with excellent versatility and generalization ability. Finally, the optimal models will be applied for icing detection in the real world.

2) Multi-step accumulation strategy: After offline training phase, the optimal MCRNN models can be obtained for real-time blade icing detection of wind turbine. To implement the real-time icing detection, a sliding window is utilized. The length of the sliding window is the same as that used in the offline training model. The sliding window moves on the available real-time data and then the trained model provides the predicted probability of icing on the each window-size-length data. However, the variable operating conditions of wind turbine and external environment factors may cause monitoring signals change a great deal, which bring about

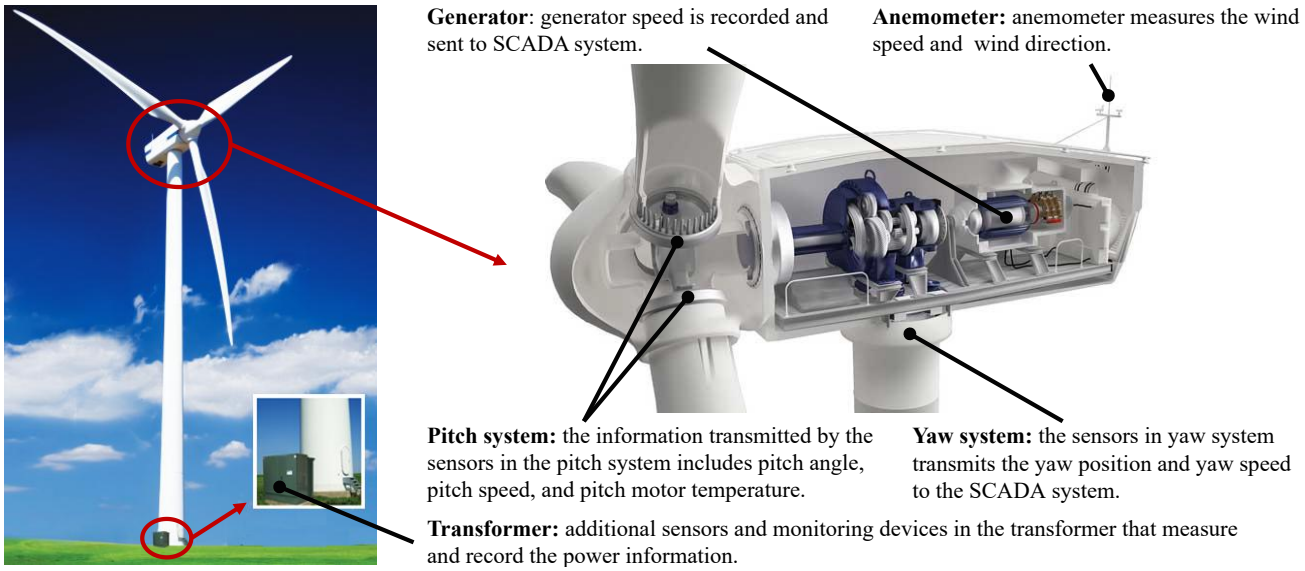

Fig. 4. Sketch of sensor location in the components of a wind turbine (modified figures from [30] and [31]). The specifications of the sensors are presented in TABLE I.

more detection errors of MCRNN in real circumstances. It will waste the energy of the monitoring personnel and make the situation more complicated if the icing detection has high false alarm rate. Therefore, a multi-step accumulation strategy is proposed to improve the accuracy of ice detection and reduce the false alarm rate. As shown in Fig. 3, assuming the predicted probability of blade icing is expressed as $[p_1, p_2, \cdots, p_{n-1}, p_n]$ For the predicted probability p_m $(m < n)$, we compare it with the threshold T. At the same time, we also need to check the comparison results of the predicted probability and the threshold of $p_{m+1} \sim p_{m+k-1}$. If k consecutive predicted probabilities are higher than the set threshold, an icing warning should be output. k can be set according to the actual application and the length of the data fragment.

E. Loss function

The raw data collected from SCADA system are characterized by severe imbalance, and therefore, the proposed data resampling method on raw data and the loss-function-based method for model optimization have been investigated. For the data resampling method, the raw data are subjected to resample to obtain balanced data sets. The cross entropy loss function is used to optimize the model when the balanced data sets are used to train the model. For the loss-function-based method, the imbalanced data sets are applied for model training directly and the focal loss function is leveraged for model optimization [32].

The cross-entropy loss (CL) function is defined as:

$$CL = -ylog(\hat{y}) + (1-y)log(1-\hat{y}) \tag{7}$$

where y represents the label of data, and the positive/negative samples are denoted as 1/0. $\hat{y}$ denotes the probability that the sample is predicted to be positive.

Focal loss (FL) [33] adds a modulating factor to the cross-entropy loss to reduce the relative loss for well-classified samples and focus on difficult samples. The equation of FL is defined as:

$$p_t = \begin{cases} \hat{y}, \, if \; y = 1, \\ 1 - \hat{y}, \, otherwise. \end{cases} \tag{8}$$
$$FL(p_t) = -\alpha_t (1-p_t)^{\gamma} log(p_t)$$

where α_t is the weight parameter between the categories used to balance the importance of positive/negative classes.

TABLE I
SPECIFICATION OF THE SCADA DATA

Number	Variable name	Description
1	wind_speed	Wind speed
2	generator_speed	Generator speed
3	power	Active power
4	wind_direction	Wind direction
5	wind_mean	Average wind direction angle within 25s
6	yaw_position	Yaw position
7	yaw_speed	Yaw speed
8	pitch1_angle	Angle of pitch 1
9	pitch2_angle	Angle of pitch 2
10	pitch3_angle	Angle of pitch 3
11	pitch1_speed	Speed of pitch 1
12	pitch2_speed	Speed of pitch 2
13	pitch3_speed	Speed of pitch 3
14	pitch1_moto_tmp	Temperature of pitch motor 1
15	pitch2_moto_tmp	Temperature of pitch motor 2
16	pitch3_moto_tmp	Temperature of pitch motor 3
17	acceleration_x	Horizontal acceleration
18	acceleration_y	Vertical acceleration
19	environment_temp	Environment temperature
20	internal_temp	Internal temperature of nacelle
21	pitch_1_ng5_temp	Switching temperature of pitch 1
22	pitch_2_ng5_temp	Switching temperature of pitch 2
23	pitch_3_ng5_temp	Switching temperature of pitch 3
24	pitch_1_ng5_DC	DC power of pitch 1 switch charger
25	pitch_2_ng5_DC	DC power of pitch 2 switch charger
26	pitch_3_ng5_DC	DC power of pitch 3 switch charger

$(1-p_t)^\gamma$ represents the modulating factor of the simple/difficult samples where γ represents the focusing parameter (simple samples belong to the category with a large number of samples while the difficult samples belong to the category with a small number of samples). In the model optimization process, the simple samples are down weighted to make the loss function pay more attention to the training of difficult samples. Consequently, employing the focal loss function to optimize the model can relieve the problem caused by imbalanced data sets that lead to the models trained on such data to perform poorly for weakly represented class (the class that has a much lower number of data points).

III. Experiments

A. Experimental setup

In this section, the experiment setup is first introduced; this mainly includes the data source and evaluation metrics. Second, the data analysis and processing are explained. Third, the performance of the proposed MCRNN model is evaluated with detailed experiments. The discussion of the experimental results is presented at the end of this section.

1) SCADA data: The experimental data are derived from standard data sets provided by the Ministry of Industry and Information Technology of the People's Republic of China, including the monitoring data of two turbines manufactured by Goldwind, Inc. and located in Inner Mongolia, China. We logged the SCADA data with a running time of 695.59 hours and 305.77 hours for these two wind turbines, respectively. The SCADA system collects data from hundreds of different sensors. According to the specific domain knowledge, engineers in the related filed have identified 26 variables related to frozen blades. These variables are specified in detail in Table I, and some of the sensor locations in the components of a wind turbine are shown in Fig. 4. The datasets from these two wind turbines are named WT-1 and WT-2.

2) Evaluation metrics: In the testing phases, consistently imbalanced data sets are used to evaluate the effectiveness of the proposed model. For the experiments on imbalanced data sets, the model can easily obtain high accuracy that is inappropriate for model evaluation, because the amount of the normal data is much higher than that of the abnormal data so that the result must be biased toward the category with a large number of the data. Therefore, **precision**, **recall** and **F1-score** are employed to evaluate the performance of the proposed model. These metrics are defined as:

$$Precision = \frac{TP}{TP+FP} \tag{9}$$

$$Recall = \frac{TP}{TP+FN} \tag{10}$$

$$F1-score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{11}$$

where TP, FP, FN and TN are true positive, false positive, false negative and true negative, respectively. Focusing on the precision, we may fail to detect the potential icing failure. By contrast, a number of false positives may be received when we merely pay attention to the recall. F1-score increase the balance of the precision and recall and thus is used as the main evaluation metric in the experiments.

TABLE II
Parameters Settings

Setting description	Setting or value
MDWD Level	3
Number of LSTM layer in LSTM branch	1
Number of hidden units in LSTM layer	16
p in Dropout layer	0.5
Number of basic convolution block in CNN branch	3
The filter sizes of the convolutional layers	64, 128, 64
The kernel sizes in convolutional layers	8, 5, 3
α_t in FL function	0.25
γ in FL function	3

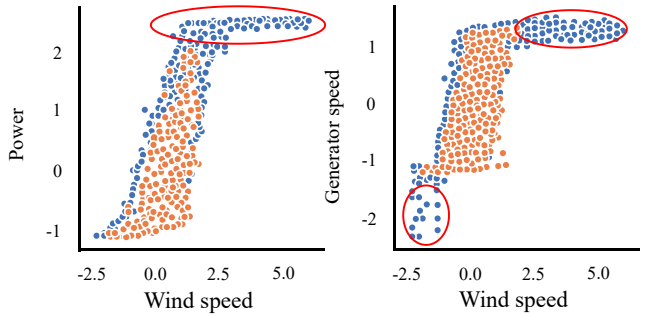

Fig. 5. Data distribution of wind speed versus power, and wind speed versus generator speed. (Blue dots represent normal data, and orange dots represent icing data). The visualized data are normalized due to the confidentiality of the original data, resulting in the loss of data units.

3) Parameter settings: In MCRNN, the MDWD level is set up to 3 which is demonstrated in Section III-E. For the LSTM branch, we set up a LSTM layer with 16 hidden units, and p is set as 0.5 in Dropout layer. Three basic blocks are set up in the CNN branch with the filter sizes of 64, 128, 64, respectively. Correspondingly, the convolution operation is carried out by three 1-D kernels with the sizes of 8, 5, 3 without striding. The step m of multi-step accumulation strategy is set as 3 during real-time icing detection phase. α_t and γ in FL function are set up to 0.25 and 3, respectively. The settings are summarized in TABEL II.

B. Data analysis and processing

1) Data analysis: The electrical machinery that generates power from wind forms the core of a wind turbine. Icing will affect the aerodynamic shape of the blade, thereby reducing the power generated at the given wind speed. The relationship between power and wind speed observed from Fig. 5 shows that for a given speed, the power of the wind turbine blades drops significantly after ice accumulates compared to the normal (non-icing) generated power. It is worth noting that the raw data is transformed to protect the sensor data privacy. Similar to power, the effect of ice accumulation on the blades also has an effect of reducing the generator speed. As the wind speed increases, the power of the wind turbine will show almost no increase after reaching the maximum value. For this time, there are obviously no data on blade ice accumulation.

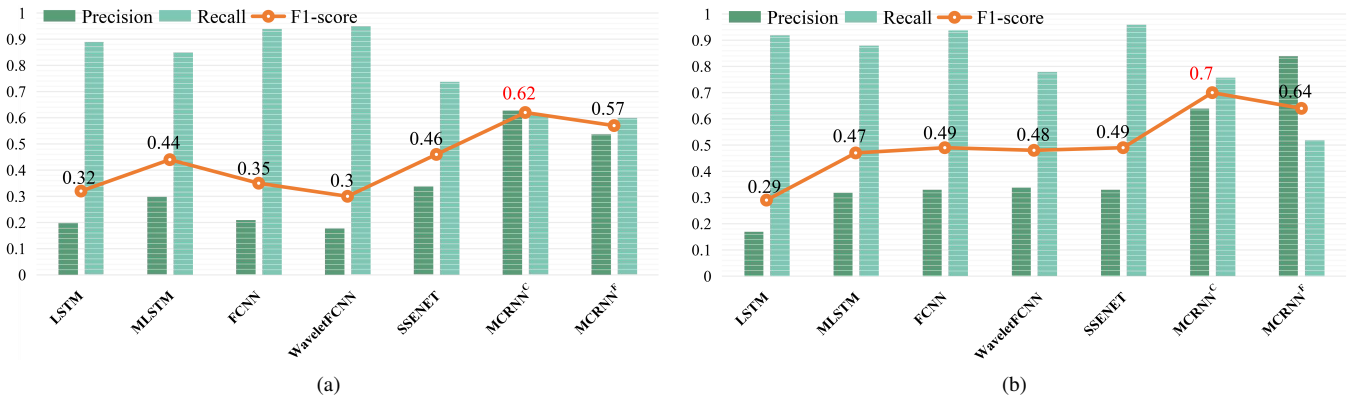

Fig. 6. Baseline comparison on balanced data sets of (a) WT-1, and (b) WT-2.

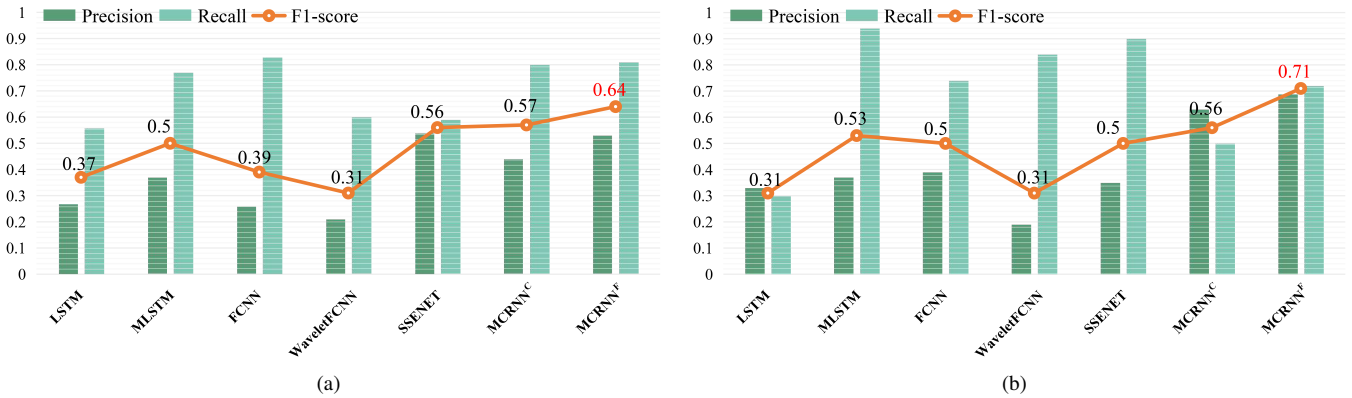

Fig. 7. Baseline comparison on imbalanced data sets of (a) WT-1, and (b) WT-2.

These redundant data (wind speed>2, generator speed<1.5) have no effect on the analysis of the blade ice accumulation. Therefore, we discard these data because they can lead to the poor feature representation ability.

Mechanically, the three blades of a normal wind turbine are balanced with strict weight and torque to avoid fatigue wear due to long-term imbalance. However, in the presence of ice accumulation, the blades are in an uncertain state, and the degree of three-leaf icing cannot be absolutely the same. Ice cubes attached to the blades may cause the wind wheel to be imbalanced with regard to weight and torque. However, examination of the relationship of the variables related to the three blades shows that even in the presence of ice accumulation, some properties of the three blades maintain good consistency. To alleviate the over-fitting of the model, the average value of the variables with the same attributes of the three blades listed in TABLE I was used as a variable for the experiment.

2) Data processing: The raw signals are split into a collection of fragments of the fixed-length time steps to use as input into the proposed model. The fixed-length time step is called as window size hereafter. Each fragment is combined with a binary label indicating whether or not the blades are frozen during this period. To solve the problem caused by data imbalance, for the data aspect, we augment the number of positive samples (samples representing icing in sensor data) by generating overlapping abnormal fragments and normal fragments without overlap regions (see Fig. 2 (a) for an example). Finally, we obtain two kind of data sets according to whether the data are resampled, where the first kind of data set is balanced, and the second is imbalanced.

C. Baseline comparison

We compare our model with four baselines on balanced data sets and imbalanced data sets; the following baseline models are used:

- **LSTM** is a time cyclic neural network, specifically designed to solve the long-term dependency problem of general RNN (recurrent neural network).
- **MLSTM FCN** is investigated for various multivariate time series classification tasks, and can be quickly deployed in real-time systems and embedded systems because of its small size and efficient characteristics [34].
- **FCNN** is a fully convolutional neural network for time series classification [35].
- **WaveletFCNN** is a deep learning modeling method for wind turbine blade icing detection [23].
- **SSENET** is a strong baseline for multivariate time series classification. SSENET utilizes the densely connection to build deep neural network. In this work, the attention mechanism used in SSENET are removed since the comparison is to verify the performance of the model structure [18].

TABLE III
COMPARISON OF DIFFERENT MODEL VARIANTS ON BALANCED DATA SETS

		MCRNN-MR^C	MCRNN-MC^C	MCRNN-CR^C	MCRNN^C	MCRNN-MR^F	MCRNN-MC^F	MCRNN-CR^F	MCRNN^F
WT-1	*Precision*	0.29	0.33	0.33	0.63	0.29	0.31	0.42	0.54
	Recall	0.91	0.79	0.75	0.61	0.84	0.80	0.58	0.60
	F1-score	0.44	0.46	0.46	**0.62**	0.43	0.45	0.49	**0.57**
WT-2	*Precision*	0.27	0.31	0.30	0.64	0.30	0.44	0.37	0.84
	Recall	0.88	0.86	0.90	0.76	0.94	0.52	0.64	0.52
	F1-score	0.41	0.45	0.45	**0.70**	0.45	0.48	0.47	**0.64**

TABLE IV
COMPARISON OF DIFFERENT MODEL VARIANTS ON IMBALANCED DATA SETS

		MCRNN-MR^C	MCRNN-MC^C	MCRNN-CR^C	MCRNN^C	MCRNN-MR^F	MCRNN-MC^F	MCRNN-CR^F	MCRNN^F
WT-1	*Precision*	0.33	0.30	0.20	0.44	0.29	0.29	0.24	0.53
	Recall	0.30	0.66	0.90	0.80	0.84	0.76	0.85	0.81
	F1-score	0.31	0.41	0.32	**0.57**	0.43	0.43	0.37	**0.64**
WT-2	*Precision*	0.29	0.43	0.18	0.63	0.44	0.56	0.43	0.69
	Recall	0.76	0.40	0.58	0.50	0.52	0.60	0.40	0.72
	F1-score	0.42	0.42	0.28	**0.56**	0.48	0.57	0.42	**0.71**

In this comparison, the window sizes of the two data sets, WT-1 and WT-2, are set to 32 and 64, respectively. In addition, MCRNN^C and MCRNN^F are two basic variants of the proposed model, where MCRNN^C uses the cross-entropy loss function during model training, and MCRNN^F uses the focal loss function, and these are applied in the two kinds of data sets: balanced data sets and imbalanced data sets.

The comparison results on balanced data sets are presented in Fig. 6. As shown in Fig. 6 (a), the F1-score improvements of MCRNN^C on WT-1 over the best baseline, SSENET, is 34.8%, and in Fig. 6 (b), the F1-score improvement over the best baseline, SSENET, on WT-2 is 42.9%. The F1-score improvements of MCRNN^F on WT-1 and WT-2 over the best baseline, SSENET, are 23.9% and 30.6%, respectively. From Fig. 6, we can observe that the result obtained by using the cross-entropy loss function is better than that obtained when the focal loss function is utilized, indicating that the rebalancing effect of the focal loss function on these balanced multivariate time series data is not competitive. Moreover, it is clear that the baseline models can achieve a high recall but a low precision, indicating that the results contain many false positives, so that most of the normal operation time of the machine is classified as icing time. These models are not suitable for real icing detection because they may cause a high false alarm rate leading to high manpower consumption to check the real operating state of the machine.

The comparison results on imbalanced data sets are presented in Fig. 7. On both data sets WT-1 and WT-2, the proposed models shows the highest F1-score. Specifically, MCRNN^C improves by 1.8% and 5.7% over the best baseline, SSENET, on both data sets, respectively, while the improvements of MCRNN^F are 14.3% and 34%, respectively. In addition, the best model for icing detection on the imbalanced data sets is MCRNN^F, and the F1-score results of MCRNN^F are 12.3% and 26.8% higher than those of MCRNN^C on the two data sets separately, indicating that the use of the rebalanced loss function, namely, focal loss function, can alleviate the imbalance of the data that gives rise to a severe result bias to the majority class.

D. Comparison to different model variants

Comparisons of three model variants are performed in order to investigate how each component of MCRNN affect its performance.

- **MCRNN-MR**: MCRNN-MR is MCRNN with MDWD and LSTM module but no CNN component.
- **MCRNN-MC**: MCRNN-MC is MCRNN with MDWD and CNN module but no LSTM component.
- **MCRNN-CR**: MCRNN-CR is MCRNN with MDWD module been removed.

The comparison results of different model variants on balanced data sets are illustrated in TABLE III, where the best scores are highlighted in bold. The results show that regardless which module in the proposed model is removed, the best performance cannot be achieved. Taking MCRNN-CR as an example, according to TABLE III, the F1-scores of MCRNN^C and MCRNN^F are 34.8% and 16.3% higher than MCRNN-CR^C and MCRNN-CR^F on WT-1 data respectively, 55.6% and 36.2% higher on WT-2 data. As shown in TABLE IV, the F1-scores of MCRNN^C and MCRNN^F are 78.1% and 73% higher than MCRNN-CR^C and MCRNN-CR^F on WT-1 data respectively, 100% and 69% higher on WT-2 data. Therefore, we can't achieve the best result with removing the MDWD module. The effect of LSTM and CNN can also be demonstrated according TABLE III and TABLE IV. From overall view, the F1-scores of MCRNN^C are 36.8% and 60.6% higher on average than those of the model variants on WT-1 and WT-2 respectively, and the results of MCRNN^F are 25.2% and 37.2% higher on average than those of the model variants on the two data sets. These results show that removal

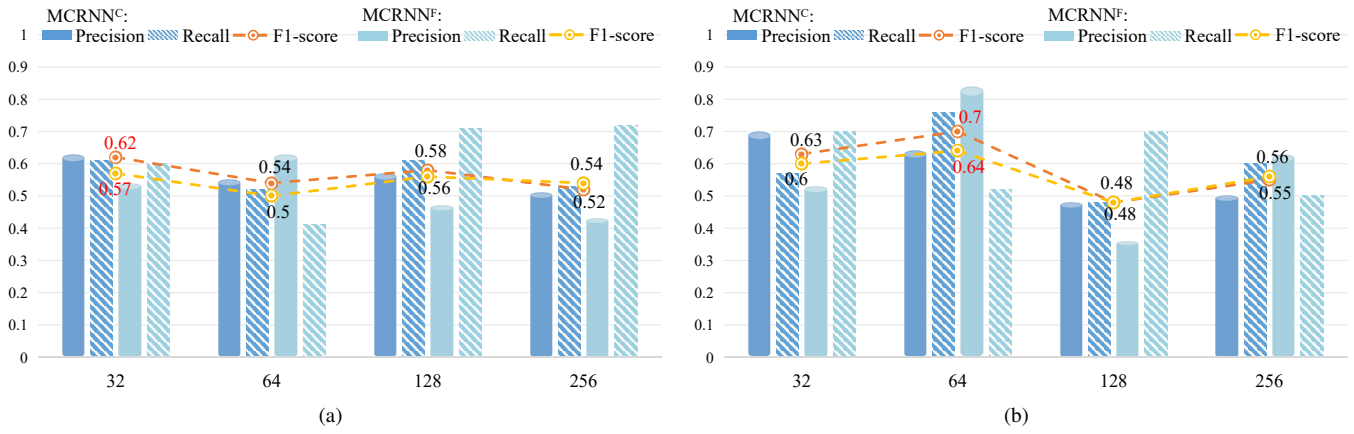

Fig. 8. Comparison on balanced data sets of (a) WT-1, and (b) WT-2 with different time steps.

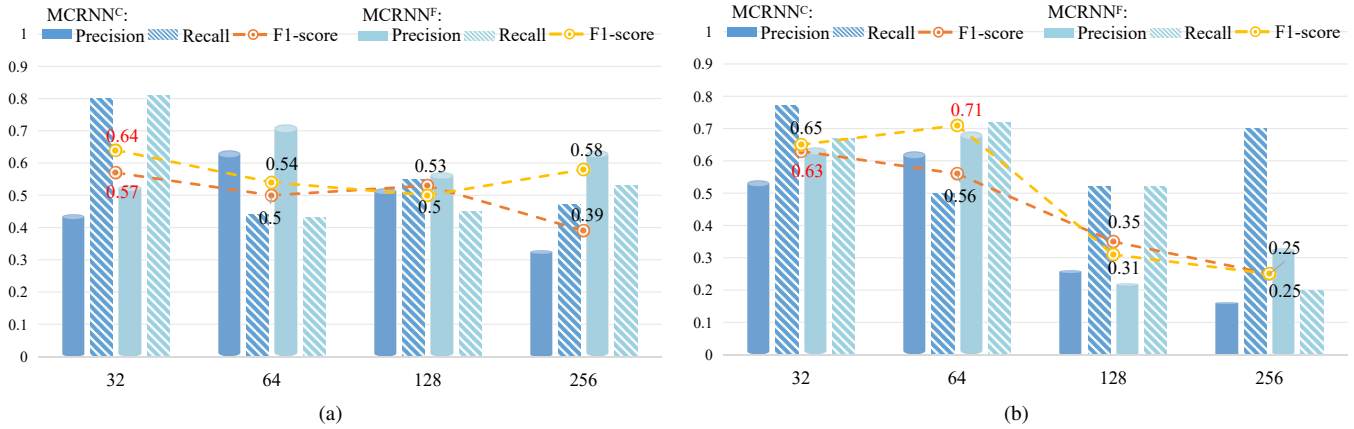

Fig. 9. Comparison on imbalanced data sets of (a) WT-1, and (b) WT-2 with different time steps.

of the LSTM branch will reduce the performance of the model, indicating the datasets have temporal dependency. Removal of the CNN branch also lead to the inability to obtain better results, proving the powerful feature extraction performance of CNN on the multivariate time series.

The comparison results of different model variants on imbalanced data sets are listed in TABLE IV. Similar to the results of comparison on balanced data sets, removal of any part of the model will reduce the model performance. In addition, all of the model variants using the focal loss function obtain higher F1-score than those using the cross entropy loss function, suggesting that the focal loss function is superior to the cross entropy loss function for the blade icing detection on imbalanced data.

E. Sensitivity analysis

Two parts of sensitivity analysis are given below. 1) Sensitivity analysis of MDWD level; 2) Sensitivity analysis of window size.

1) Impact of different MDWD level: Intuitively, the analysis results of MDWD with disparate levels on sequences may be different. Thus sensitivity analysis of MDWD level is conducted in order to verify the influence of different MDWD levels on icing detection results and to find which is the best level we should set up in icing detection. The decomposition is conducted on low-frequency components that generated from the 1/2 down-sampling of the intermediate variable sequences of each level, so it will result in low frequency resolution if the sequence of the last level of decomposition is too short. Therefore, according to the length of the data segment, we conduct experimental analysis on the four level of 2, 3, 4, and 5.

In the experiments, balanced data sets are employed com-

TABLE V
MDWD LEVEL SENSITIVE ANALYSIS

		MDWD Level			
		2	3	4	5
WT-1 Balanced data & CL function	*Precision*	0.45	0.63	0.45	0.41
	Recall	0.72	0.61	0.80	0.71
	F1-score	0.55	**0.62**	0.59	0.52
WT-1 Imbalanced data & FL function	*Precision*	0.62	0.53	0.58	0.65
	Recall	0.64	0.81	0.64	0.52
	F1-score	0.63	**0.64**	0.61	0.59
WT-2 Balanced data & CL function	*Precision*	0.49	0.64	0.55	0.49
	Recall	0.86	0.76	0.82	0.78
	F1-score	0.62	**0.7**	0.66	0.6
WT-2 Imbalanced data & FL function	*Precision*	0.56	0.69	0.66	0.43
	Recall	0.66	0.72	0.42	0.84
	F1-score	0.61	**0.71**	0.51	0.57

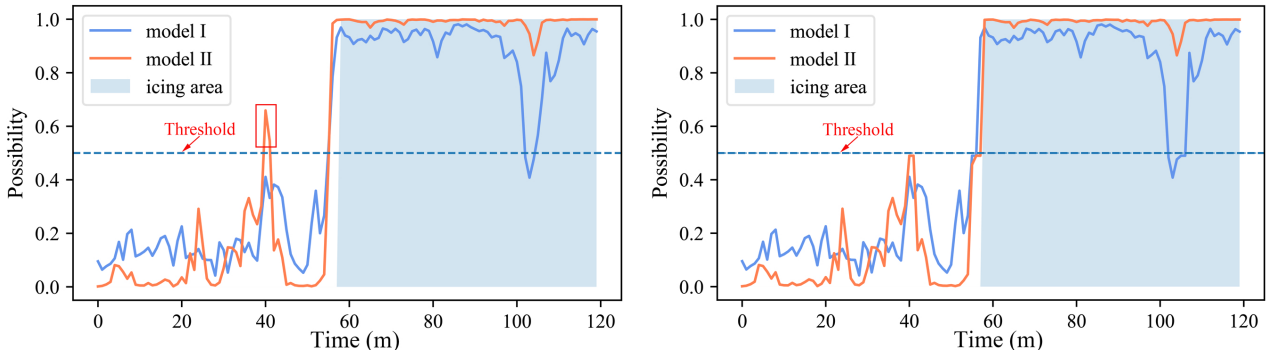

Fig. 10. Real-time detection on WT-1 data with (a) directly detection and (b) detection with multi-step accumulation strategy. 'm' indicates minute, which means icing detection is conducted every minute. Model I and II indicate the optimal models (listed in TABLE VI) employed for real-time icing detection, and the information of the models is listed in TABLE VI.

TABLE VI
INFORMATION OF OPTIMAL MODELS

Model	Training data	Model	F1-score	Precision	Recall
I	WT-1	MCRNN^C	0.62	0.63	0.61
II	WT-1	MCRNN^F	0.64	0.53	0.81

bining with cross-entropy loss function, and imbalanced data sets combining with focal loss function. The results of sensitivity analysis of MDWD level are illustrated in TABLE V with the highest F1-score highlighted in bold. It can be known from TABLE V that the highest F1-score can be obtained by all the models when the wavelet transform level is 3, which means that using wavelet transform level of 3 can get better time domain and frequency domain information for feature extraction than the others.

2) Impact of different window size: To investigate the influence of window size in data segmentation, sensitivity analysis of different window size is conducted. The sensitivity analysis is conducted on two data sets: balanced data set and imbalanced data set. In this comparison, four window size of 32, 64, 128 and 256 are set for data segmentation in both data sets, corresponding to approximately 3.5, 7, 15 and 30 minutes of data, respectively.

The comparison results of balanced data set are shown in Fig. 8. The comparison results on WT-1 show that the F1-score obtained using the window size of 32 is higher than the values obtained using other window size values, while the results on WT-2 indicate that the window size of 64 is optimal. All of the data contained in a segment are generated within a few minutes regardless of whether this segment uses 32 or 64 as the window size for the data segmentation. The advantage of using such a shorter segmentation length is that icing can be detected in a shorter and earlier time. As shown in Fig. 9 (a) and Fig. 9 (b) for the comparison results of the imbalanced data set, the results on WT-1 and WT-2 demonstrates that step sizes of 32 and 64 are superior to the other two step sizes, same as found in the balanced data set.

F. Real-time icing detection

In order to further verify the practical application of the proposed model, real-time icing detection is performed. Real-time icing detection is conducted on part of WT-1 test dataset. As mentioned in Section II, the data sets are split into fixed-length data segments used for model training, so the same length data segment should be collected each time for real-time detection. A sliding window of fixed-length moves forward on the available sensors data, and the stride is sets to 8 corresponding to about one minute. Thus, icing detection is done every minute. The information of optimal models generated in the training phase for icing detection is listed in TABLE VI. The probability threshold of icing and normal is set to 0.5 which is of an average level. In the real-use scenario, the threshold should be further determined by observed relationship between the output probability and ice amount in the blades. In order to improve the robustness of the detection method in this paper and keep the false alarm rate as low as possible, the proposed multi-step accumulation strategy is applied in real-time detection. When the probability of icing prediction is greater than the threshold for less than k consecutive times, the probability will be reduced to less than the threshold. k is set to 3 here.

The results are illustrated in Fig. 10. From of the visualization graphs, Fig. 10 (a) shows the results of direct real-time detection, while Fig. 10 (b) shows the results of real-time detection with multi-step accumulation strategy. From an overall point of view, the two model training methods (using balanced data with cross-entropy loss function and using imbalanced data with focal loss loss function) both can detect all icing areas. From the perspective of false alarm rate, as show in Fig. 10 (a), there are false alarms (curve in the red box) in the non-icing periods by carrying out real-time icing detection with direct alarm. As can be seen in Fig. 10 (b), employing multi-step accumulation alarm strategy can obviously reduce these false alarms, thereby effectively reducing the non-icing periods false alarm rate.

G. Discussion

By conducting experiments on balanced data sets and imbalanced data sets, we find that MCRNN outperforms the state-

of-the-art approaches in terms of F1-score in Figs. 6 and 7, demonstrating the effectiveness and superiority on multisensor time series data for icing detection task. In view of the severe imbalance characteristic of the data generated by SCADA system, we consider the problem for the data aspect and algorithm aspect, respectively. For the data aspect, the highest F1-score is achieved by MCRNN^C on the balanced data sets, and the F1-scores achieved by MCRNN^C are 16.9% higher on average than the results on the no-sampling data sets. The effect of the data aspect solution suggests that conducting resampling on imbalanced data is favorable for the capture of the correlation on multivariate time series by the proposed model. From the algorithm aspect, MCRNN^F can achieve the best performance on the imbalanced data sets. In addition, the solution for algorithm aspect can achieve $1.4\% \sim 3.2\%$ higher F1-score than the solution for the data aspect. Therefore, our experiments verify the effectiveness of the two solutions for addressing the imbalanced data, particularly the effect of using the focal loss function during the model training phase on imbalanced multivariate time series data. In addition, we can find the best MDWD level is 3 for the proposed MCRNN to achieve best performance. Moreover, the results in the sensitivity analysis demonstrate that a shorter time step can provide better icing detection results, which is more relevant with real needs. Early activation of the deicing system or the shutdown of the turbine for maintenance is beneficial for machine maintenance and energy loss reduction. Furthermore, the results of real-time detection prove that the multi-step accumulation strategy can reduce the false alarm rate, so in real application scenarios, leveraging the strategy can improve the robustness of icing detection models and bring about reduction of the energy consumption of monitoring personnel.

IV. Conclusion

In this paper, with the goal of meeting the challenges in detection of icing of wind turbine blades, we propose MCRNN that can learn the multiple channel correlations and multiple scale dynamic patterns in the analysis of time series data. The model contains two basic modules. 1) The first is a data transformation module, where multilevel discrete wavelet decomposition is leveraged to obtain the time domain and frequency domain information of the data that are beneficial and crucial for the feature learning. 2) In addition, we effectively make use of LSTM to analyze the temporal dependency of the data, and the powerful feature extraction capability of CNN is also used. Thus the proposed model can maximize the leverage of time-frequency domain information of the original data and the data transformed by MDWD. To address the severe imbalance of data, we propose two solutions from the data and algorithm aspects. Furthermore, multi-step accumulation strategy is proposed to enhance the robustness of the proposed MCRNN to detect icing on blades. A series of experimental comparisons fully verify the generalization performance of the proposed MCRNN model, the effectiveness of the solution on imbalanced data and the necessity of multi-step accumulation strategy. The results demonstrate that the proposed model outperforms several methods for wind turbine blades icing detection.

Acknowledgment

This work was supported by National Natural Science Foundation of China (62020106004, 92048301). We especially thank GoldWind, Inc. for sharing the data.

References

[1] Z. Ren, A. S. Verma, Y. Li, J. J. Teuwen, and Z. Jiang, "Offshore wind turbine operations and maintenance: A state-of-the-art review," *Renewable and Sustainable Energy Reviews*, vol. 144, p. 110886, 2021.

[2] "Top 10 wind turbine manufacturers in the world." https://www.bizvibe.com/blog/top-10-wind-turbine-manufacturers-world/.Accessed: 2019-02-03.

[3] K. Wei, Y. Yang, H. Zuo, and D. Zhong, "A review on ice detection technology and ice elimination technology for wind turbine," *Wind Energy*, vol. 23, 2020.

[4] E. Madi, K. Pope, W. Huang, and T. Iqbal, "A review of integrating ice detection and mitigation for wind turbine blades," *Renewable and Sustainable Energy Reviews*, vol. 103, pp. 269–281, 2019.

[5] P. Wang, W. Zhou, Y. Bao, and H. Li, "Ice monitoring of a full-scale wind turbine blade using ultrasonic guided waves under varying temperature conditions," *Structural Control & Health Monitoring*, vol. 25, no. 4, pp. e2138.1–e2138.17, 2018.

[6] V. Carlsson, "Measuring routines of ice accretion for wind turbine applications : The correlation of production losses and detection of ice," Master's thesis, Umeå University, 2010.

[7] X. Bin, G. Xueyi, and C. Hongbing, "Active icing monitoring for wind turbine blade models with pzt technology," *Piezoelectrics & Acoustooptics*, vol. 39, p. 02, 2017.

[8] P. Roberge, J. Lemay, J. Ruel, and A. Bégin-Drolet, "A new atmospheric icing detector based on thermally heated cylindrical probes for wind turbine applications," *Cold Regions Science and Technology*, vol. 148, pp. 131–141, 2018.

[9] D. K. Hsu, F. J. Margetan, S. J. Wormley, and J. A. Simpson, "Apparatus and method for detection of icing onset and ice thickness," Mar. 17 1992, uS Patent 5,095,754.

[10] T. Laakso, H. Holttinen, G. Ronsten, L. Tallhaug, R. Horbaty, I. Baring-Gould, A. Lacroix, E. Peltola, and B. Tammelin, "State-of-the-art of wind energy in cold climates," *IEA annex XIX*, vol. 24, p. 53, 2003.

[11] M. L. Corradini, A. Cristofaro, and S. Pettinari, "A model-based robust icing detection and estimation scheme for wind turbines," in *2016 European Control Conference (ECC)*. IEEE, 2016, pp. 1451–1456.

[12] Q. Shi, "Model-based detection for ice on wind turbine blades," Master's thesis, NTNU, 2017.

[13] L. Hu, X. Zhu, J. Chen, X. Shen, and Z. Du, "Numerical simulation of rime ice on nrel phase vi blade," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 178, pp. 57–68, 2018.

[14] P. Song, Z. Yao, and C. Zhao, "Section division and multi-model method for early detection of icing on wind turbine blades," in *2019 34rd Youth Academic Annual Conference of Chinese Association of Automation (YAC)*. IEEE, 2019, pp. 749–754.

[15] G. A. Skrimpas, K. Kleani, N. Mijatovic, C. W. Sweeney, B. B. Jensen, and J. Holboell, "Detection of icing on wind turbine blades by means of vibration and power curve analysis," *Wind Energy*, vol. 19, no. 10, pp. 1819–1832, 2016.

[16] Z. Guangfei, T. Wen, and Z. Da, "Ice detection for wind turbine blades based on pso-svm method," *Journal of Physics: Conference Series*, vol. 1087, p. 022036, 09 2018.

[17] H. Yi, Q. Jiang, X. Yan, and B. Wang, "Imbalanced classification based on minority clustering smote with wind turbine fault detection application," *IEEE Transactions on Industrial Informatics*, 2020.

[18] X. Cheng, G. Li, A. L. Ellefsen, S. Chen, H. P. Hildre, and H. Zhang, "A novel densely connected convolutional neural network for sea-state estimation using ship motion data," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 9, pp. 5984–5993, 2020.

[19] A. L. Ellefsen, V. Æsøy, S. Ushakov, and H. Zhang, "A comprehensive survey of prognostics and health management based on deep learning for autonomous ships," *IEEE Transactions on Reliability*, vol. 68, no. 2, pp. 720–740, 2019.

[20] L. Chen, G. Xu, Q. Zhang, and X. Zhang, "Learning deep representation of imbalanced scada data for fault detection of wind turbines," *Measurement*, vol. 139, pp. 370 – 379, 2019.

[21] Y. Liu, H. Cheng, X. Kong, Q. Wang, and H. Cui, "Intelligent wind turbine blade icing detection using supervisory control and data acquisition data and ensemble deep learning," *Energy Science & Engineering*, vol. 7, no. 6, pp. 2633–2645, 2019.

[22] X. Cheng, F. Shi, M. Zhao, G. Li, H. Zhang, and S. Chen, "Temporal attention convolutional neural network for estimation of icing probability on wind turbine blades," *IEEE Transactions on Industrial Electronics*, pp. 1–1, 2021.

[23] B. Yuan, C. Wang, F. Jiang, M. Long, P. S. Yu, and Y. Liu, "Waveletfcnn: A deep time series classification model for wind turbine blade icing detection," *ArXiv*, vol. abs/1902.05625, 2019.

[24] Y. Cui, M. Jia, T.-Y. Lin, Y. Song, and S. Belongie, "Class-balanced loss based on effective number of samples," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 9268–9277.

[25] X. Wang, G. Tang, T. Wang, X. Zhang, B. Peng, L. Dou, and Y. He, "Lkurtogram guided adaptive empirical wavelet transform and purified instantaneous energy operation for fault diagnosis of wind turbine bearing," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–19, 2020.

[26] J. Wang, Z. Wang, J. Li, and J. Wu, "Multilevel wavelet decomposition network for interpretable time series analysis," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018, pp. 2437–2446.

[27] K. Kawakami, "Supervised sequence labelling with recurrent neural networks," *Ph. D. dissertation, PhD thesis. Ph. D. thesis*, 2008.

[28] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *Proceedings of the 32nd International Conference on Machine Learning, PMLR 37:448-456*, 2015.

[29] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, 2010, pp. 807–814.

[30] "Wind turbine step-up transformers." https://studylib.net/doc/18480053/wind-turbine-step-up-transformers.Accessed:2020-09-03.

[31] "Predictive maintenance of wind turbines and generators." https://www.maintenanceandengineering.com/2019/08/12/predictive-maintenance-of-wind-turbines-and-generators/.Accessed: 2020-09-03.

[32] P. Han, G. Li, R. Skulstad, S. Skjong, and H. Zhang, "A deep learning approach to detect and isolate thruster failures for dynamically positioned vessels using motion data," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.

[33] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[34] F. Karim, S. Majumdar, H. Darabi, and S. Harford, "Multivariate lstm-fcns for time series classification," *Neural Networks*, pp. 237–245, 2018.

[35] Z. Wang, W. Yan, and T. Oates, "Time series classification from scratch with deep neural networks: A strong baseline," in *2017 International Joint Conference on Neural Networks (IJCNN)*, 2017, pp. 1578–1585.

Weiwei Tian is currently pursuing the master degree with School of Computer Science and Technology, Tianjin University of Technology, Tianjin, China. Her current research interests include artificial intelligence, deep learning, decision support, predictive maintenance, and digital twins.

Xu Cheng received the Ph.D. degree from the Department of Ocean Operations and Civil Engineering, Norwegian University of Science and Technology, Aalesund, Norway, in 2020. Xu Cheng has published more than 20 papers as first and coauthor in the areas of his research interests which include data analytics and artificial intelligence in maritime operations, time series analytics, and predictive maintenance of wind turbines.

Guoyuan Li (M'14-SM'19) received the Ph.D. degree in computer science from the Department of Informatics, Institute of Technical Aspects of Multimodal Systems, University of Hamburg, Hamburg, Germany, in 2013. In 2014, he joined the Department of Ocean Operations and Civil Engineering, Intelligent Systems Laboratory, Norwegian University of Science and Technology (NTNU), Ålesund, Norway. He is currently a Professor of ship intelligence at NTNU. He has published more than 70 articles in the areas of his research interests, including modeling and simulation of ship motion, autonomous navigation, intelligent control, optimization algorithms, and locomotion control of bioinspired robots.

Fan Shi received a Ph.D. degree from Nankai University, Tianjin, China, in 2012. He is currently an associate professor at Tianjin University of Technology, China. From June 2018 to August 2019, he was a research scholar in West Virginia University. His research interests include machine vision, pattern recognition and optics.

Shengyong Chen (SM'10) received the Ph.D. degree in computer vision from the City University of Hong Kong, Hong Kong, in 2003. He was with the University of Hamburg from 2006 to 2007. He is currently a Professor with the Tianjin University of Technology, China. He received the Fellowship from the Alexander von Humboldt Foundation of Germany. He worked as a visiting professor at Imperial College London and University of Cambridge, U.K. He has over 100 patents and published over 400 scientific papers in international journals and conferences, with 50 in IEEE Transactions. His research interests include autonomous robots, computer vision, and machine intelligence. Selected publications can be found at his website http://sychen.com. Dr. Chen is an IET Fellow and an IEEE senior member. He received the National Outstanding Youth Foundation Award of China in 2013. He received the IEEE Sensors Journal Best Paper Award in 2012 and IET Premium Award in 2013. He organized over 20 international conferences and serves as associate editors of three international journals, e.g. IEEE Trans. Cybernetics.

Houxiang Zhang is a full Professor at the Department of Ocean Operations and Civil Engineering, Faculty of Engineering, Norwegian University of Science and Technology (NTNU).

Dr. Zhang received his Ph.D. degree on Mechanical and Electronic Engineering in 2003. From 2004, he worked as Postdoctoral fellow, senior researcher at the Institute of Technical Aspects of Multimodal Systems (TAMS), Department of Informatics, Faculty of Mathematics, Informatics and Natural Sciences, University of Hamburg, Germany. In Feb. 2011, he finished the Habilitation on Informatics at University of Hamburg. Dr. Zhang joined the NTNU, Norway in April 2011 where he is a Professor on Mechatronics. From 2011 to 2016, Dr. Zhang also hold a Norwegian national GIFT Professorship on product and system design funded by Norwegian Maritime Centre of Expertise. In 2019, Dr. Zhang has been elected to the member of Norwegian Academy of Technological Sciences.

Dr. Zhang has engaged into two main research areas including control, optimization and AI application especially on autonomous vehicle; and marine automation, digitalization and ship intelligence. He has applied for and coordinated more than 30 projects supported by Norwegian Research Council (NFR), German Research Council (DFG), EU, and industry. In these areas, he has published over 200 journal and conference papers as author or co-author. Dr. Zhang has received four best paper awards, and five finalist awards for best conference paper at International conference on Robotics and 'Automation'.